

В. Шаряк

*Івано-Франківський інститут менеджменту і економіки
„Галицька академія”*

МЕТОДИ ДОСЛІДЖЕННЯ СИСТЕМНИХ ХАРАКТЕРИСТИК МОДЕЛЕЙ БАЗИ ДАНИХ

У статті розглянуто типи бази даних, проведено дослідження системних характеристик моделей: ієрархічної, мережевої, реляційної, лінійно-рекурентної, об'єктно-орієнтованої, об'єктно-реляційної та напівструктурованої бази даних. Визначено оцінку та зменшення надлишковості, яка розрахована на основі критерію емерджентності (інтелектуальності в залежності від числа зв'язків та числа елементів бази даних. Визначено критерії ефективності на основі емерджентності, надлишковості, швидкості пошуку даних і методи оптимізації й підвищення ефективності бази даних і на перспективи розвитку архітектури бази даних.

V. Sharyak

METHODS OF RESEARCHES OF DESCRIPTIONS OF SYSTEMS OF MODELS OF DATABASE

The types of database are considered in the article, researches of descriptions of the systems of models: of hierarchical, network, relational, linearly-recurrent, object-oriented, object-relational, semistructured database. An estimation and diminishing of it is surplus is certain expected on the basis of criterion of emergence (to intellectuality) in dependence of numbers of connections and numbers of elements of database. The criteria of efficiency are certain on the basis of emergence, to surplus, speeds of retrieval of data and methods of optimization and increase of efficiency of database and prospect of development of architecture database.

Умовні позначення

БД – бази даних;

СКБД – системи керування базами даних;

Ne – критерій емерджентності (інтелектуальності);

Ke – коефіцієнт критерію емерджентності (інтелектуальності);

N₃ – число зв'язків,

N₀ – число елементів БД;

mG – об'єм кодової інформації для ідентифікації всіх груп елементів;

mR – кодова розрядність ідентифікатора в базисі Радемахера;

mGn – ефективність кодування БД в базисі Галуа шляхом кодування числа елементів у групі після коду ідентифікатора;

ΔN_i – число тактів ідентифікації даних;

T – час доступу до даних.

Вступ

Сучасні технології БД є одним із визначних факторів вдосконалення процесів зберігання та доступу до великих масивів даних. Потужність сучасних БД базується на результатах досліджень та технологічних розробках, які отримані протягом кількох останніх десятиріч. Управління БД виконується спеціалізованими програмами (СУБД), які є ефективним інструментом підтримки БД. Особливим видом БД є розподілені БД, важливими характеристиками яких є ступінь паралелізму операцій та швидкого пошуку даних. Слід зауважити, що технології розподіленої обробки інформації ускладнюють всі компоненти і спонукають переглядати їх структури аж до базових елементів. При цьому в багатьох розподілених БД затрати на підтримку телекомунікацій можуть переважити вартість обчислювальних процесів та програмно-технічного забезпечення.

Тому при дослідженні та проектуванні розподілених БД найважливішу роль відіграють задачі компактного кодування даних та швидкості обміну даними, при заданій пропускій здатності каналного зв'язку. Отже, задача вдосконалення архітектури та методів кодування інформації в розподілених БД є актуальною [1,2,3,4,5,6]

1. Типи моделей БД

Розрізняють сім типів моделей баз даних : 1) ієрархічні; 2) мережеві; 3) реляційні; 4) лінійно – рекурентні; 5) об'єктно – орієнтовані; 6) об'єктно – реляційні; 7) напівструктуровані.

1.1 Ієрархічні БД

Ієрархічна структура даних - це структура, де будь-який об'єкт може підпорядковуватися лише одному об'єкту вищого рівня, а йому багато об'єктів нижчого рівня. Такий зв'язок між даними називають „один до багатьох”. За ієрархічним принципом побудовані файлові структури даних на дисках персонального комп'ютера [1,2]. На рис.1 показані зв'язки ієрархічної моделі БД.

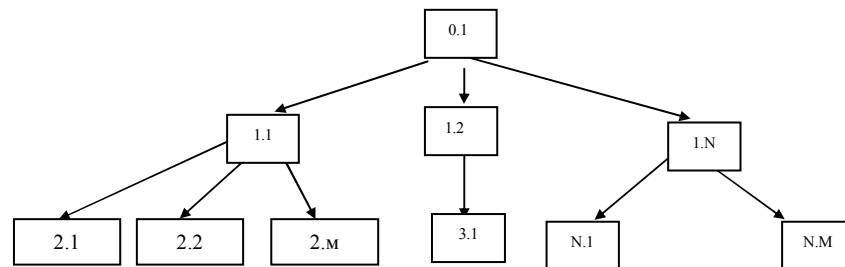


Рисунок 1 - Ієрархічна модель БД

1.2 Мережева БД

Мережева структура характеризується тим, що будь-який об'єкт одного рівня (одної групи даних) може мати довільні зв'язки з об'єктами іншого рівня. Такі зв'язки називаються „багато до багатьох”. Мережеві структури можна описати у вигляді таблиці, де у першому горизонтальному рядку записують об'єкти одного рівня, а у першому вертикальному – іншого. Така таблиця добре ілюструє зв'язки між об'єктами, але може мати багато порожніх елементів, що призводить до значної надлишковості БД [1,2]. На рис.2 показані зв'язки мережевої моделі БД.

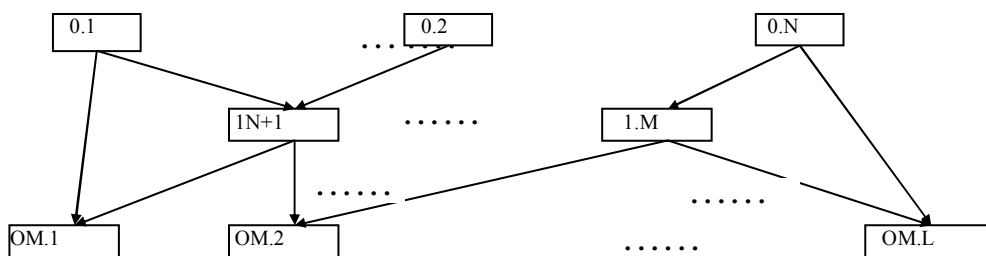


Рисунок 2 - Мережева модель БД

1.3 Реляційні БД

Реляційна БД є найпоширенішим типом і характеризується поданням даних у вигляді декількох таблиць і зв'язаними між собою кортежів та атрибутів. Один з найпростіших типів зв'язків є „один до одного”. Потрібні дані будуть черпатися з двох таблиць. Отже, для зв'язку між таблицями використовують поле, значення якого не повторюється в різних записах. Це поле називається ключовим. Якщо реляційні таблиці мають спільні поля, то зміни у спільному полі в одній таблиці автоматично відображаються у всіх таблицях. Мета запровадження реляційних зв'язків – мінімізувати дублювання даних і забезпечити можливість опрацювати (шукати) дані з

декількох таблиць, що забезпечує значне зменшення надлишковості БД [1,2]. На рис.3 показана структура реляційної моделі БД.

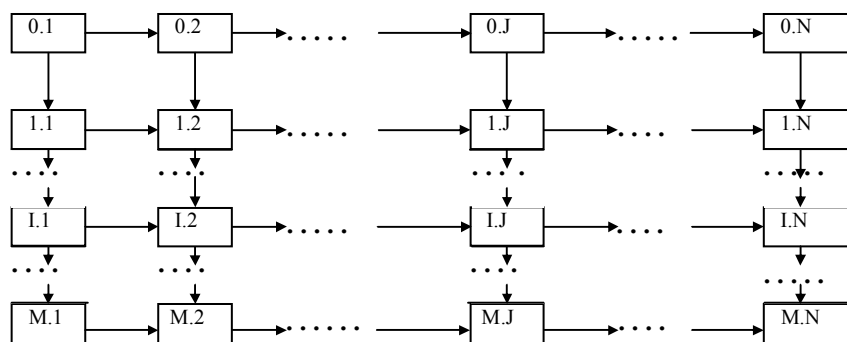


Рисунок 3 - Реляційна модель БД

1.4 Лінійно-рекурентні БД

Лінійно-рекурентні БД є новим типом структурної організації баз даних, яка формується на основі рекурентних властивостей теоретико-числового базису Галуа. Важливою характеристикою таких БД є максимальна компактність адресації даних, в тому числі при багаторівневій архітектурі. Позитивною характеристикою лінійно-рекурентних БД є можливості ефективного захисту від помилок та несанкціонованого доступу [4].

1.5 Об'єктно-орієнтовані БД

Один із підходів до забезпечення сумісності систем БД з парадигмою об'єктно-орієнтованого проектування зв'язаний з розширенням системи розуміння, на основі об'єктно – орієнтованих мов програмування, таких як C++ або Java. В традиційному програмуванні розуміється, що після завершення циклу роботи програми її об'єкти безповоротно втрачаються, в той час як принципові умови роботи будь-якої СУБД полягають у тому, що об'єкти повинні зберігатися необмежено довгий час, поки не будуть змінені або знищені примусово, як це відбувається у файлових системах. „Чисту” об'єктно-орієнтовану модель представлення даних, названу *ODL (Object Definition Language)* - мова визначення об'єктів, яка в свій час була стандартизована дослідницькою групою Object Data Management Group (ODMG) [1]. На рис.4 показана структура об'єктно-орієнтованої БД.

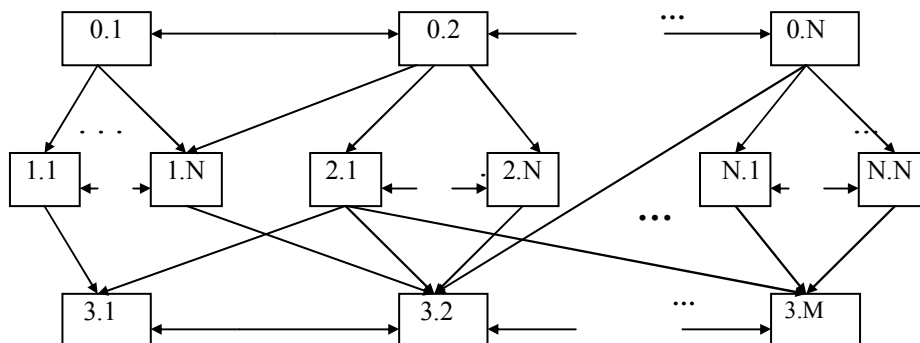


Рисунок 4 - Об'єктно – орієнтовані БД

1.6 Об'єктно-реляційні БД

Ця модель є частиною самого останнього стандарту SQL, названого SQL-99 (або SQL-1999, також SQL3), і являє собою варіант розширення звичайної реляційної моделі за рахунок формалізації багатьох загальноприйнятих концепцій об'єктно-орієнтованого проектування. Вказаний стандарт служить основою для побудови об'єктно-реляційних систем БД, які сьогодні випускаються всіма основними постачальниками комерційних СУБД. Ці системи, однак, значно відрізняються в

деталіях практичної реалізації початкових концепцій [1]. На рис.5 показана структура об'єктно-реляційної БД.

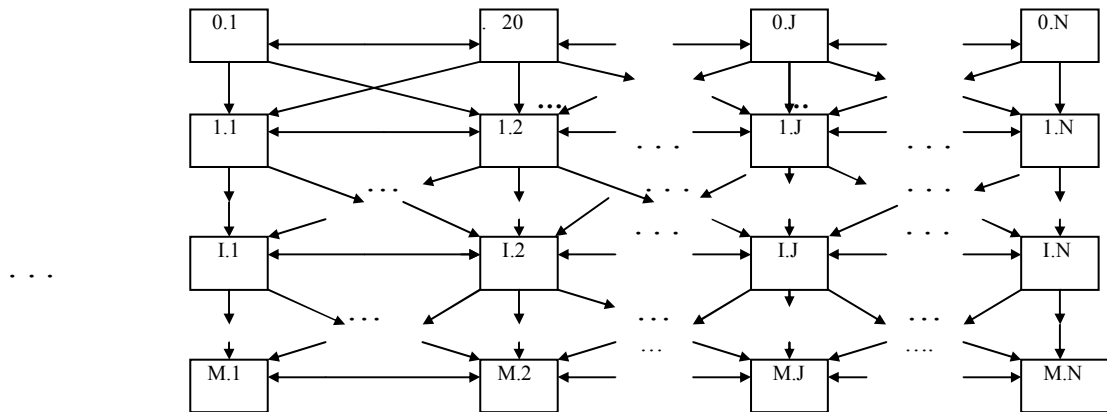


Рисунок 5 - Об'єктно – реляційна модель БД

1.7 Напівструктуровані БД

Одне із останніх досягнень, призначених вирішити велику кількість актуальних проблем технологій СУБД, включаючи необхідність об'єднання традиційних баз даних з іншими джерелами інформації, такими, як Web-сторінки найрізноманітнішої структури. В той час, як об'єктно-орієнтовані або об'єктно-реляційні системи пропонують використання фіксованих схем для кожного класу або відношень, модель напівструктурованих даних відрізняється набагато більш високою гнучкістю представлення компонентів інформації. Найбільш яскравим практичним втіленням моделі напівструктурованих даних є XML (eXtensible Markup Language-розширювана мова розмітки). XML-це специфікація опису „документів”, що представляють собою набір вложений елементів даних, функції яких визначаються відповідними тегамі. Підтримка XML з часом стане найважливішим компонентом систем, що здійснюють посередницькі функції між різнорідними джерелами даних, і навіть, що дуже можливо, послужить цілям гнучкого представлення інформації в БД.[1].

2. Оцінка надлишковості БД

Зменшення надлишковості в різних моделях БД досягається за рахунок зростання числа зв'язків між об'єктами БД. Таким чином, оцінку надлишковості моделі БД можна розрахувати на основі критерію емерджентності (інтелектуальності) [1,2.6]

$$K_e = \frac{N_3}{N_0}, \quad (1)$$

де N_3 -число зв'язків, N_0 -число елементів БД.

В [2] показано, що система об'єктів (наприклад: даних) характеризується емерджентністю в тому випадку, коли

$$K_e \geq 2. \quad (2)$$

Результати дослідження емерджентності різних типів БД наведені в таблиці 1, звідки видно, що ієрархічні структури ніколи не забезпечують умови емерджентності.

3. Критерії ефективності БД

Важливими характеристиками ефективності БД є:

- емерджентність (1);

- надлишковість

$$K_e = \frac{mR}{mG}; \quad (3)$$

-швидкість пошуку даних

$$V = \frac{\Delta N_i}{T}, \quad (4)$$

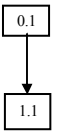
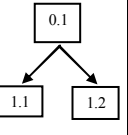
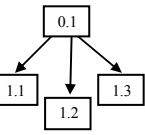
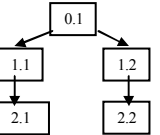

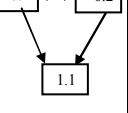
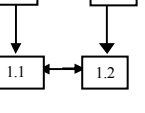
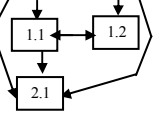
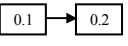
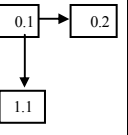
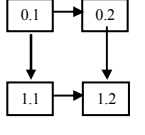
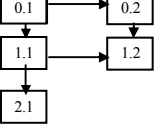
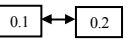

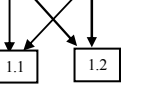
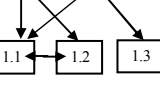
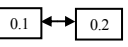
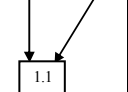
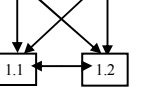
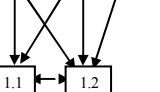
де: mR - ентропія ієрархічних даних в базисі Радемахера;
 mG – ентропія рекурентних даних в базисі Галуа;
 ΔN_i - число тактів ідентифікації даних;
 T - час доступу до даних.

4. Фактори які впливають на підвищення ефективності БД

Основними характеристиками оптимізації і підвищення ефективності є :

- здешевлення системи СУБД ;
- збільшення швидкості пошуку доступу до даних;
- зменшення або стиснення об'єму інформації ;
- розробка різноманітних типів третинних пристроїв зберігання (tertiary storage devises) даних;
- тенденції росту системи ;
- підвищення продуктивності системи ;
- паралельні обчислення;
- розвиток системи клієнт/сервер і багаторівневої архітектури;
- підвищення надійності захисту від помилок;
- захист від несанкціонованого доступу до БД тощо [1,2].

Таблиця1 - Результати досліджень емерджентості різних типів БД

Тип БД	$N_0=2$	$N_0=3$	$N_0=4$	$N_0=5$	$N_0=100$
Ієрархічна	$K_e = \frac{1}{2} = 0.5$ 	$K_e = \frac{2}{3} = 0.66$ 	$K_e = \frac{3}{4} = 0.75$ 	$K_e = \frac{4}{5} = 0.8$ 	... $K_e < 1$
Мережева	$K_e = \frac{2}{2} = 1.0$ 	$K_e = \frac{4}{3} = 1.33$ 	$K_e = \frac{6}{4} = 1.5$ 	$K_e = \frac{9}{5} = 1.8$ 	... $K_e \geq 2$
Реляційна	$K_e = \frac{1}{2} = 0.5$ 	$K_e = \frac{2}{3} = 0.6$ 	$K_e = \frac{4}{4} = 1.0$ 	$K_e = \frac{5}{5} = 1.0$ 	... $K_e < 2$
Об'єктно-орієнтована	$K_e = \frac{2}{2} = 1.0$ 	$K_e = \frac{4}{3} = 1.33$ 	$K_e = \frac{5}{4} = 1.5$ 	$K_e = \frac{8}{5} = 1.6$ 	... $K_e \leq 4$
Об'єктно-реляційна	$K_e = \frac{2}{2} = 1.0$ 	$K_e = \frac{4}{3} = 1.33$ 	$K_e = \frac{8}{4} = 2.0$ 	$K_e = \frac{11}{5} = 2.2$ 	... $K_e < 4$

Висновки

Проведені дослідження архітектури БД показують, що ієрархічна, мережеві та реляційні архітектури БД не задовільняють умови емерджентності. Тому перспективним є розвиток та досконалення архітектури БД об'єктно-орієнтованих та об'єктно-реляційних архітектур.

Література

1. В.В.Шаряк, Я.М.Николайчук. Системні характеристики баз даних та перспективні напрямки їх розвитку // Моделювання та інформаційні технології. -Київ. ІЕД НАН України.№1(1), 2007.-С.38-44.
2. Шаряк В.В., Николайчук Я.М. Теоретико-числові бази та їх застосування при організації бази даних // Оптико-електронні інформаційноенергетичні технології.-ВНТУ.Вінниця.№2(12), 2006. - С.59-66.
3. Шаряк В. Архітектура і кодування баз даних на основі теоретико-числових базисів. //Вісник. ТДТУ.- Тернопіль, №1, (3) 2007.- С.17-23.
4. Дж. Мартин. Введение в организацию баз данных . -М.:Мир,1980. -662 с.
5. Гарсиа-Молина и другие. Системы баз данных. -М.: Вильямс , 2004. –1088с.
6. Пасічник В.В.,Резніченко В.А. Організація баз даних та знань. -К.: Видавнича група ВНУ, 2006. - 384с.

Одержано 05.03.2008 р.